Introduction to Health Data Analytics for Clinicians

A short course that provides a high-level introduction to the world of health data research

The course is designed for hospital and primary care physicians, nurses, pharmacists, allied health providers, and other health professionals with an interest in leveraging data to address clinically-relevant questions. On completion of the course, participants will have an understanding of the considerations involved in planning and leading an analytic study and the tools available to support cutting edge analysis.

The course is offered collaboratively by the Melbourne School of Population and Global Health and the Health and Biomedical Informatics Centre at the University of Melbourne. The course will feature expert speakers in informatics, machine learning, and health services research and will provide a forum for interaction, discussion and collaboration with like-minded peers.

Participants will generally be novices in the field of data analytics but will be required to have a strong grasp of general research concepts.

**This course will be held at Graduate House
220 Leicester Street, Carlton**

Early bird fees – until 28 February 2019: $825 (GST inclusive)

Normal rates (non-early bird): $935 (GST inclusive)

(Light working supper provided)

The course will be delivered over three evenings

6:30pm to 9:30pm on:
**Wednesday 17 April 2019
Wednesday 8 May 2019
Wednesday 29 May 2019**

🖥 **Registration:** http://go.unimelb.edu.au/3xq6
✉ **Enquiries:** HDA-enquiries@unimelb.edu.au

Course Outline

The course is delivered over three evenings. The outline for each day is provided below.

Participants will be asked to identify their key objectives in undertaking the short course in order to tailor the content to the group's learning needs. Opportunities for further, formal study at the University will be highlighted at the end of each session. The in-house nature of the course is designed to facilitate relationships with session leaders and peer research collaborators.

Module 1: Planning and Designing a Study

In this module you will be introduced to the key concepts involved in defining an analytic research question and designing an efficient study. We will explore important considerations involved in analytic projects, including ethics and risk management principles.

Module 2: Mapping and Accessing Data

In this module we will provide an overview of the types of data that can be used for health research and how they can be accessed. You will learn how to critically appraise available data resources and scope the data required to answer a question of interest.

Module 3: Analytical Approaches

In this module we will provide an overview of the types of statistical techniques commonly applied in analytic projects. You will be introduced to data-driven analytic techniques such as machine learning and gain an understanding of the tools and resources available to support computationally intensive analysis.

Certificates of attendance will be provided and CME accreditation for major professional colleges (RACP, RACS, ANZCA, ACEM, CICM) is being sought

Introduction to Health Data Analytics for Clinicians

New Non- Award Course - Course Outline

Graeme Hart: HaBIC and Ximena Camacho: Melbourne School of Population and Global Health

Intended Audience

- Early to mid-career clinicians
- Senior clinical / professorial staff with research experience
- Clinicians with an interest in data analytics and visualization
- Target sectors: hospital, primary care, nursing, pharmacy, allied health

Proposed Mode of Delivery

- Three, 3-hour lectures in the evenings (6:30 – 9:30pm)
- Class size: 20-25 participants
- Location
    o Lectures delivered on the University campus, open to anyone interested
- May need to be offered several times per year

Course Framework

One of the aims of the course is to develop a community of practice among the clinical community.

Pre-course survey

Data to be collected via LMS as part of pre-workshop

- Participant introduction (who you are / background)
- Why do you want to participate in the course
- What are your interests and expectations for the course

The course content has been consolidated around three broad thematic areas.

Theme 1: Planning a study / contextualizing the data required

The project planning methods that mid- or senior-career clinicians might use when developing a research study may be different to those used when developing a randomized controlled trial. A clinical question will define the data elements required for a research project, which will need to be sourced from existing collections. Methods for defining control groups or matching criteria may differ based on the data available compared to those applied when collecting data.

Proposed lecture breakdown:

- Welcome, review of objectives, review of course outline

Module 1 Thematic discussion

- Overview of results of pre-course poll (i.e. what is everyone interested in)

Getting started

- Finding a home for your research (overview of potential journals)
    - Research can also be targeted at relevant non-medical journals (e.g. policy, public health, biostats, informatics)
    - Important to start all research with a literature review
        - Replication in a different jurisdiction (cross-comparisons)
        - quickly and easily identifies the data you need, methodology
        - critically appraise existing publications, how would you address your critics (think about this ahead of time and use this to help you design your study)
    - Choose a target journal / field (this will help focus the research)
- Define the research question
    - Identify the study population, study period, exposure, outcome, covariates
- Identify the analytic approach
    - Bring in subject matter experts to support the project (e.g. (bio)statisticians, health economists, clinical collaborators, informaticians, analysts)
- Outline the proposed article
    - Use this as a framework for the study methodology and data requirements
    - Outline required tables/figures to focus analysis
    - Identify control / comparison groups and matching criteria
    - Review STROBE guidelines
- Communicate the results
    - Concepts around communicating with non-clinical / non-academic audiences (e.g. policy makers, health planners, patients, etc.)

Research governance - Assoc Prof Mark Taylor University of Melbourne Law School

- Ethics
    - HRECs
    - Consent
    - Direct care delivery / interventions
- Multi-sectoral linkages
    - Data custodians, linkage units, integrating authorities
- 5 safes
- Consumer involvement / patient perspectives

(*provide references / resources relevant to theme discussion*)

Theme 2: Mapping and accessing the data

- Types of data
    - Quantitative vs qualitative

- o Structured vs unstructured data
- o Administrative vs survey vs registry vs other
- o "Omics" data
- Overview of available repositories
  - o How to find out what data is available and where it exists
- Data collection
  - o Data dictionaries, metadata
  - o Terminology, coding / classification systems and mapping
    - https://www.digitalhealth.gov.au/get-started-with-digital-health/what-is-digital-health/clinical-terminology
    - https://www.healthterminologies.gov.au/
    - LOINC, SNOMED, drug terminology (e.g. ATC / drug class, generic status, brand, drug name, etc.)
  - o Versions (ICD-9 vs ICD-10); updates to coding systems within a version; updates to coding practices/definitions over time
- Local data linkage assets
  - o Dr Anthony Carpenter – School of Population and Global Health
  - o Dr Jane Hunter- Director Australian Urban Infrastructure Network
  - o CVDL (admin data linkage)
  - o Polar (GP)
  - o GRHANITE (GP)
  - o DARE (Hospitals Clinical Research Data Warehouse)
  - o BioGrid
  - o Others (e.g. ABS, AIHW; - omics)
  - o AURIN
    - Social and Urban Infrastructure, geospatial mapping tools / social determinants data, use cases of AURIN health related research, collaboration and research networks, research consent frameworks
- Data access process
  - o Approvals
    - Ethics
    - Linkage rules
    - Cross-jurisdictional collaborations
  - o Clinical collaborators (in hospital)
  - o BioGrid governance framework
  - o Privacy
    - Reporting size (small cells), geographical identifiers, rare diseases/events
  - o Security
    - Secure access, data retention, destruction
- Mapping data required for a project
  - o Identifying the data sources you need
  - o What if the data you want doesn't exist?
    - How to collect additional data elements required (in the context of an EMR or other survey tool)
      - REDCap, EMR, SMART on FHIR
- Next steps – further study (e.g. graduate certificates, graduate degrees), accredited courses, university/hospital resources (e.g. MCATS, DARE), fellowship program

Module 2 Thematic discussion (15-30 mins)

- Follow up on online discussions from Module 1
    - Participants to come prepared for discussion
    - *references / resources relevant to discussion*

Theme 3: analytical approaches   - Dr Christopher McMaster

- Data linkage methodologies
    - What is data linkage and why do we need it?
    - Different types of linkage (probabilistic vs deterministic)
    - What it can do / what it can't do
- Data cleaning
    - Key risks (missing, dates, etc.) – high level
- Statistical approaches
    - Overview of techniques commonly used in data analysis projects (e.g. regression, survival analysis, competing risks, propensity scores [adjustment / matching])
    - Software / tools (e.g. SAS, R, RStudio, Stata, etc.)
        - Have a general sense of what the techniques are and when they would be applied, but should really bring a statistician on board to undertake / oversee this part
- AI approaches
    - Overview of machine learning techniques (e.g. neural networks, random forests, NLP, etc.)
    - Platforms, publicly available tools (e.g. Watson, Deep 6 AI, Microsoft Azure platform)
- Visualization tools
- Computing needs
    - Many machine learning and statistical approaches are computationally intensive; will need to have appropriate resources / computing power
    - Overview of resources available
        - UoM (e.g. NECTAR, high performance computing, etc.)
        - Other platforms (e.g. Watson, Deep 6 AI, Microsoft Azure platform, cloud services)
    - Other "non-traditional" tools (e.g. NVivo, etc.)
        - Integration / application to qualitative research
- Next steps – further study (e.g. graduate certificates, graduate degrees), accredited courses, university/hospital resources (e.g. MCATS, DARE), fellowship program (*provide references / resources relevant to theme discussion*)

Module 3 Thematic discussion (15-30 mins)

- Real-life project examples
    - Datathon?
- Follow up on online discussions from Module 2
    - Participants to come prepared for discussion

(*references / resources relevant to theme discussion*)